INTERNATIONAL JOURNAL OF TRANSGENDER HEALTH
2021, VOL. 22, NO. 4, 360–380
https://doi.org/10.1080/26895269.2021.1915224

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE



# Effects of speech therapy for transgender women: A systematic review

Clara Leyns[a], Tine Papeleu[a], Peter Tomassen[b], Guy T'Sjoen[c,d] and Evelien D'haeseleer[a,e]

[a]Center for Speech and Language Sciences (CESLAS), Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium; [b]Department of Head and Neck Surgery, Ghent University Hospital, Ghent, Belgium; [c]Department of Endocrinology, Ghent University Hospital, Ghent, Belgium; [d]Center for Sexology and Gender, Ghent University Hospital, Ghent, Belgium; [e]Department of Otorhinolaryngology, Ghent University Hospital, Ghent, Belgium

**ABSTRACT**

***Background:*** For transgender women, communication and speech characteristics might not be congruent with their gender expressions. This can have a major influence on their psychosocial functioning. Higher quality of life scores were observed the more their voice was perceived as feminine. Speech language pathologists may play an important role in this, as the gender affirming hormone treatment for transgender women does not affect the voice.
***Aim:*** This systematic review aimed to provide speech and language pathologists with the current literature concerning the effects of speech therapy in transgender women in terms of acoustic and perceptual outcomes.
***Methods:*** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was used for reporting this systematic review. The Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE (using the PubMed interface) and Embase (using the embase.com interface) were used as electronic databases. All individual studies which measured the effects of speech therapy in transgender women were evaluated with a risk of bias assessment tool and levels of evidence. Relevant data were extracted from these studies and a narrative synthesis was performed.
***Results:*** 14 studies were identified through the databases and other sources. These studies show positive outcome results concerning pitch elevation, oral resonance, self-perception and listener perception. However, methodological issues contribute to problems with generalization and reproducibility of the studies.
***Conclusion:*** There is an urgent need for effectiveness studies using RCT designs, larger sample sizes, multidimensional voice assessments, well-described therapy programs, investigators blinded to study process, and longer-term follow-up data. Speech and language pathologists who work with transgender women may find these results essential for defining therapy goals.

**KEYWORDS**
Effectiveness; PRISMA; speech therapy; systematic review; transgender voice; transgender women

## Introduction

Communication and speech characteristics of transgender women might not be congruent with their gender expressions. This can have a major influence on their psychosocial functioning (Colton & Casper, 1996). A study by Hancock et al. (2011) observed higher quality of life scores of transgender women the more their voice was perceived as feminine. Speech language pathologists may play an important role in this, as the gender affirming hormone treatment for transgender women does not affect the voice (Gooren, 2005; Hancock & Garabedian, 2013).

Voice and communication training in order to feminize the voice includes altering voice characteristics such as speaking fundamental frequency, fundamental frequency range, intonation patterns, loudness, vocal quality and resonance (Dacakis, 2000). In the literature, therapy goals for voice feminization have been described based on these voice aspects that influence listener's gender perceptions. Research regarding listener's perceptions, such as the results of a systematic review and meta-analysis by Leung et al. (2018), showed that aspects which are most salient in listener's perceptions of speaker gender are primarily fundamental frequency ($f_o$) of the voice and secondly resonance characteristics. Loudness,

**CONTACT** Clara Leyns clara.leyns@ugent.be Center for Speech and Language Sciences (CESLAS), Department of Rehabilitation Sciences, Corneel Heymanslaan 10, Ghent 9000, Belgium.

articulation, and intonation were also found to be associated with listener's perceptions of speaker gender. In contrast, tempo and stress were not significantly associated and mixed results were found concerning the contribution of breathiness to gender perception. The results of this review suggested that the $f_o$ of the voice contributes for 41.6% of the variance in gender perception. Listeners' perceptions may not change from male to female or masculine to feminine by altering pitch alone. The authors of the review described resonance as the second most widely studied vocal domain concerning listener perceptions of speaker gender. Resonance depends on the length and shape of the vocal tract which can be altered to change the frequencies of the vowel formants (De Bodt et al., 2015; Meister et al., 2017). This can be done by adjusting vowel characteristics such as jaw drop or mouth opening, lip spreading, tongue position and duration of the vowel (Carew et al., 2007; Gallena et al., 2018; Timmermans, 2013). The first three characteristics play a role in adjusting the resonance of the voice and determine the frequencies of the first three formants (F1, F2, F3) (Carew et al., 2007; Cartei et al., 2012; Coleman, 1971; Corthals, 2008; De Bodt et al., 2015; Fant, 1966; Gallena et al., 2018; Günzburger, 1995; Ladefoged, 1993; Meister et al., 2017; Mount & Salmon, 1988; Nordström, 1977; Pisanski & Rendall, 2011; Titze, 1989; Weirich & Simpson, 2018; Wu & Childers, 1991). According to Corthals (2008), F1 is related to the jaw drop (larger jaw angle induces a higher F1). Mount and Salmon (1988) correlated F2 to the degree of fronting of the tongue, meaning, a more forward tongue position induces a higher F2. Günzburger (1995) and Corthals (2008) concluded that F3 increases considerably when there is less lip protrusion and when the oral cavity is shortened by elevating the tongue and larynx. As these characteristics change the vowel formants, they can influence resonance.

The vowel formant frequencies of larger vocal tract cavities are lower (Fant, 1966; Nordström, 1977; Titze, 1989; Wu & Childers, 1991). In general, the formants of cisgender men are twenty percent lower than those of cisgender women (Coleman, 1983). These differences in formant frequencies are too large to be caused by purely anatomical aspects (Günzburger, 1995). It is possible that people change their vocal characteristics to comply with either female or male speech stereotypes. Oates and Dacakis (1983) stated that on average, cisgender women articulate more precisely and accurately compared to cisgender men. A study by Cartei et al. (2012) asked 17 cisgender men and 15 cisgender women to imitate the voice of the opposite gender. These imitations were compared with the pretest measurements in terms of $f_o$, formants 1 till 4 (F1-F4) and the degree of lip spreading and mouth opening. When asked to imitate masculinity, formant frequencies dropped, inducing a smaller vowel space, and the opposite happened when imitating female voices. Furthermore, female voices showed larger lip spreading than male voices on average. Although no statistically significant differences were found, mouth opening, a key determinant of F1, was found to be larger in cisgender women compared with cisgender men. Moreover, cisgender women were found to have a larger vowel space which might be associated with a perceptual femininity of the voice (Weirich & Simpson, 2018).

A listening experiment conducted by Pisanski and Rendall (2011) observed that listeners use formant characteristics as a cue to gender attribution. The review by Leung et al. (2018) showed that the frequency of the first 4 formants (F1-F4) contribute to gender perception, i.e. higher formant frequencies contribute to a more female gender perception. The lower the first formant frequency, the more masculine a voice is judged (Weirich & Simpson, 2018). Gallena et al. (2018) investigated gender perception of the voice after increasing both $f_o$ and formant frequencies. If the $f_o$ is in the gender ambiguous zone (150 Hz – 185 Hz, Mordaunt, Adler & Hirsch (2006)), the voice of transgender women is nevertheless often perceived as that of cisgender men when the formant frequencies are still in the male area. A minimum increase of 20% of all formant frequencies and the resulting increase of vowel space results in a statistically significant increase of perceiving a voice as female or feminine. Therefore, it would be more effective to aim at a moderate increase of all relevant formant frequencies and vowel space (Gallena et al., 2018).

Although therapy goals can be identified based on determinants of listener's perceptions, as reviewed by Leung et al. (2018), it is not yet clear whether voice and communication training focusing on these goals is successful, i.e. transgender women sound more feminine after the intervention and are satisfied with the outcome. Research on these intervention outcomes for transgender people is limited. Preliminary results of pilot studies in transgender women are promising and suggest that voice and verbal communication training could result in vocal changes, gender perception and patient reported outcome measures (Carew et al., 2007; Dacakis, 2000; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Mészáros et al., 2005; Söderpalm et al., 2004; Van Borsel et al., 2000). As there is no evidence for effectiveness of targeting nonverbal communication with transgender women, this review focused on the verbal aspect of communication (Davies et al., 2015; Oates, 2006).

To our knowledge, this is the first systematic review that has summarized the evidence of the effects of speech therapy. In the past, reviews in this specific research area have been performed by Oates and Dacakis (1983), Dacakis et al. (2012) and Davies et al. (2015). Oates and Dacakis (1983) gave a full overview of 'the Nature of Transsexualism', 'Sex Reassignment Programmes in Melbourne', speech markers of male and female speech, speech stereotypes, and how to 'manage the communication problems of transsexuals', which included some recommendations for clinicians working in this field. However, much of the early literature focused on single-case studies or small-group studies that report the outcomes of speech therapy. A more recent review was performed by Dacakis et al. (2012), which described recent research findings that demonstrate the effectiveness of speech pathology intervention for transgender women. This review brought out the importance of modifying vocal resonance and encouraging precise articulation to increase perceptions of a feminine voice. Davies et al. (2015) collected evidence from previous studies concerning effectiveness of speech training for transgender individuals. They intended to support clinicians and researchers who are relatively new working in this field of speech language pathology, with specific recommendations concerning clinical competence, treatment decisions, assessment, client inclusion, etc. However, a review which follows a strict systematic methodology, such as using the guidelines of the preferred reporting items for systematic reviews and meta analyses statement (PRISMA; Moher et al., 2015) is very useful to guide speech language pathologists in defining therapy goals when working with transgender women. The aim of this systematic review was to provide speech and language pathologists with the current literature concerning the effects of speech therapy in transgender women in terms of acoustic and perceptual outcomes.

## Methods

The methodology and reporting were based on recommendations from the Cochrane Collaboration (Higgins et al., 2019) and the preferred reporting items for systematic reviews and meta analyses statement (PRISMA; Moher et al., 2015).

### *Protocol and registration*

This review was conducted according to the protocol previously published in the PROSPERO register (registration number CRD42020192000).

### *Eligibility criteria*

The eligibility criteria (see Table 1) were predefined to answer the research question.

### *Search*

The following electronic databases were searched from inception until July 27, 2020: the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE (using the PubMed interface) and Embase (using the embase.com interface). The cited references of the included studies were also checked. Furthermore, ProQuest Dissertations & theses global (proquest.com), Open Gray (opengrey.eu) and Open Access Theses and Dissertations (oatd.org) were searched for gray literature, including theses or dissertations, pre-prints, conference presentations, posters, abstracts or unpublished manuscripts.

The search strategy for PubMed was reported in Table 2. Two concepts were combined with a Boolean operator AND. In the other databases, equivalent search terms were used and MeSH terms were adapted to the relevant database.

**Table 1.** Eligibility criteria.

| | Inclusion | Exclusion |
|---|---|---|
| Publication type | • Randomized controlled trials, non-randomized studies (cohort studies, case control studies and cross sectional studies and quasi randomized controlled clinical trials, case reports) will be considered as appropriate study designs for inclusion.<br>• Peer-reviewed studies<br>• Published in English<br>• No limitation concerning the publication date | • Reviews, systematic reviews, meta-analyses<br>• Studies which are not peer-reviewed<br>• No full text available |
| Participants | • Transgender women: assigned male at birth and female gender identity<br>• No age restrictions | • Gender non-binary people<br>• Transgender men<br>• Transgender women who completed phonosurgery |
| Intervention | • Pitch elevation<br>• Articulation – resonance<br>• Intonation | |
| Intervention outcomes | • Fundamental frequencies ($f_0$)<br>• Fundamental frequency range<br>• Formant frequencies<br>• Vowel space<br>• Patient satisfaction<br>• Gender perception | |

Endnote X9 (Clarivate Analytics) was used for creating four separate libraries: (1) Records identified through database searching and identified through other sources (including duplicates), (2) Records after duplicates removal, (3) Records included based on title and abstract screening, and (4) Records included based on full-text evaluation.

### Study selection

Title-abstract and full-text screening was performed independently by the first two authors (CL and TP) through the Rayyan platform (Ouzzani et al., 2016). The authors determined whether the article was appropriate according to the eligibility criteria. Any discrepancies were discussed, and a final list of articles was subject to data analysis. The study selection process was reported in a flow diagram (Moher et al., 2009).

### Data collection process and data items

The reviewers were not masked to the author, institution, and publication source of trials at any time. Using piloted extraction forms two reviewers independently extracted the characteristics of the trials, baseline characteristics of the

**Table 2.** Search strategy for MEDLINE (PubMed).

| Concept | Search strategy |
|---|---|
| Concept 1: *Transgender women* | **MeSH terms:**<br>"Transgender Persons"[Mesh] OR "Transsexualism"[Mesh] OR "Gender Identity"[Mesh] OR "Gender Dysphoria"[Mesh] OR<br>**Free text words:**<br>"gender variant*"[TIAB] OR "gender-variant*"[TIAB] OR "gender queer"[TIAB] OR "gender-queer"[TIAB] OR genderqueer[TIAB] OR "gender nonconforming"[TIAB] OR "2 spirit person"[TIAB] OR "2-spirit person"[TIAB] OR "two spirit person"[TIAB] OR "two-spirit person"[TIAB] OR "male to female transsexual*"[TIAB] OR "male-to-female transsexual*"[TIAB] OR MTF[TIAB] OR M2F[TIAB] OR "M-To-F transsexual*"[TIAB] OR AMAB[TIAB] OR "assigned male at birth"[TIAB] OR "TG girl*"[TIAB] OR "male-to-female-transgender"[TIAB] OR genderidentity[TIAB] OR "gender identity"[TIAB] OR "gender-identity"[TIAB] OR genderdysphoria[TIAB] OR "gender-dysphoria"[TIAB] OR "gender dysphoria"[TIAB] OR "gender minorit*"[TIAB] OR transgend*[TIAB] OR "trans-gend*"[TIAB] OR transsex*[TIAB] OR "trans-sex*"[TIAB] OR transex*[TIAB] OR transvest*[TIAB] OR "gender non-conforming"[TIAB] OR "gender ambiguous"[TIAB] OR "gender bender"[TIAB] OR transfem*[TIAB] OR transwoman[TIAB] OR transwomen[TIAB] OR "trans woman"[TIAB] OR "trans women"[TIAB] OR "trans female"[TIAB] OR "trans feminine"[TIAB] OR "trans-feminine"[TIAB] OR transfeminine[TIAB] OR transfemale[TIAB] OR "trans people"[TIAB] OR transpeople[TIAB] OR "male to female"[TIAB] OR "sex reassign*"[TIAB] OR "sex change"[TIAB] OR "gender reassign*"[TIAB] OR "gender confirm*"[TIAB] OR "gender chang*"[TIAB] OR "gender transition"[TIAB] OR "gender disorder"[TIAB] OR GLB[TIAB] OR GLBQ[TIAB] OR GLBT[TIAB] OR GLBTQ[TIAB] OR LGB[TIAB] OR LGBT[TIAB] OR LGBQ[TIAB] OR LGBTQ[TIAB] OR "LGBTQ-people"[TIAB] OR "sexual identit*"[TIAB] OR "sexual minorit*"[TIAB] |
| Concept 2: *Speech therapy* | **MeSH terms:**<br>"Rehabilitation of Speech and Language Disorders"[Mesh] OR "Speech-Language Pathology"[Mesh] OR<br>**Free text words:**<br>voice[TIAB] OR voicetherapy[TIAB] OR "vocal therapy"[TIAB] OR "vocal rehabilitation"[TIAB] OR feminization[TIAB] OR feminization[TIAB] OR pitch[TIAB] OR intonation[TIAB] OR prosody[TIAB] OR "verbal communication"[TIAB] OR "oral communication"[TIAB] OR "verbal behavior"[TIAB] OR "verbal behavior"[TIAB] OR speech[TIAB] OR logopedic[TIAB] OR logopaedic*[TIAB] OR logopedics[TIAB] OR logopedica[TIAB] OR logotherapy[TIAB] OR logotherapies[TIAB] OR "language pathologist"[TIAB] OR "language pathologists"[TIAB] OR "speech-language"[TIAB] OR "language pathology"[TIAB] |

participants, the description of intervention and outcomes. Discrepancies were resolved through consensus. The items extracted from each article from the selection process were as follows:

1. Study design. In cases where the study design was not explicitly stated, the authors deciphered the design from the method described in the article.
2. Study population characteristics
3. Intervention characteristics
4. Outcome characteristics

### Risk of bias and levels of evidence of individual studies

To assess the risk of bias, the quality assessment tool 'QUALSYST' from the "Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields" was used (Kmet et al., 2004). With this tool, 14 items of each quantitative study were scored on the study and outcome levels depending on the degree to which the specific criteria were met or reported ("yes" = 2, "partial" = 1, "no" = 0). Items not applicable to a particular study design were marked "N/A" and were excluded from the calculation of the summary score. A percentage was calculated for each paper by dividing the total sum score obtained across rated items by the total possible score.

In order to investigate the weighting of the included studies, an adapted version of the Levels of Evidence by Sackett (1989) was used (Burns et al., 2011). The levels can be found in the appendix. The first author rated each of the articles against the risk of bias assessment tool and the levels of evidence.

## Results

### Study selection

The PRISMA Flow Diagram (Figure 1) summarizes the review process and selection of the studies meeting inclusion criteria. Overall, we retrieved 1671 records from the systematic searches in 3 databases. After exclusion of duplicates and non-relevant records, a total of 13 studies remained. 3 additional studies were checked for inclusion through other sources. This resulted in 14 articles for data extraction. Reasons for excluding references during full text screening were wrong language (n = 8), wrong study design (n = 6) and wrong population (n = 5).

### Study characteristics

The characteristics per study were presented in Table 3.

### Risk of bias assessment and levels of evidence

The risk of bias assessment total sum scores for the reviewed studies ranged from 4 to 20 with a mean of 11.79. The total summary scores, taking the total possible sum (i.e. without the N/A questions) into account, ranged 18% to 82%, with a mean score of 49%. An overview of the results can be seen on Figure 2. In Table 3, the levels of evidence can be found. 11 studies had a level V of evidence, indicating little or no systematic empirical evidence. Only Kawitzky and McAllister (2020), Gelfer and Tice (2013) and Gelfer and Van Dong (2013) had a level II, which shows that their findings are generally consistent (Burns et al., 2011).

### Study design

Most studies (n = 11) were prospective study designs, only Dacakis (2000), Hancock and Garabedian (2013) and Söderpalm et al. (2004) reported a retrospective design. Gelfer and Tice (2013) Gelfer and Van Dong (2013) used cis male and female control participants to provide speech samples for acoustic analysis, and Kawitzky and McAllister (2020) included cis male control participants for their intervention as well. Four studies included control speech samples for their listening experiment (Bralley et al., 1978; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Kaye et al., 1993).

### Study population characteristics

Across all studies in this systematic review, 95 transgender women were included. Their mean

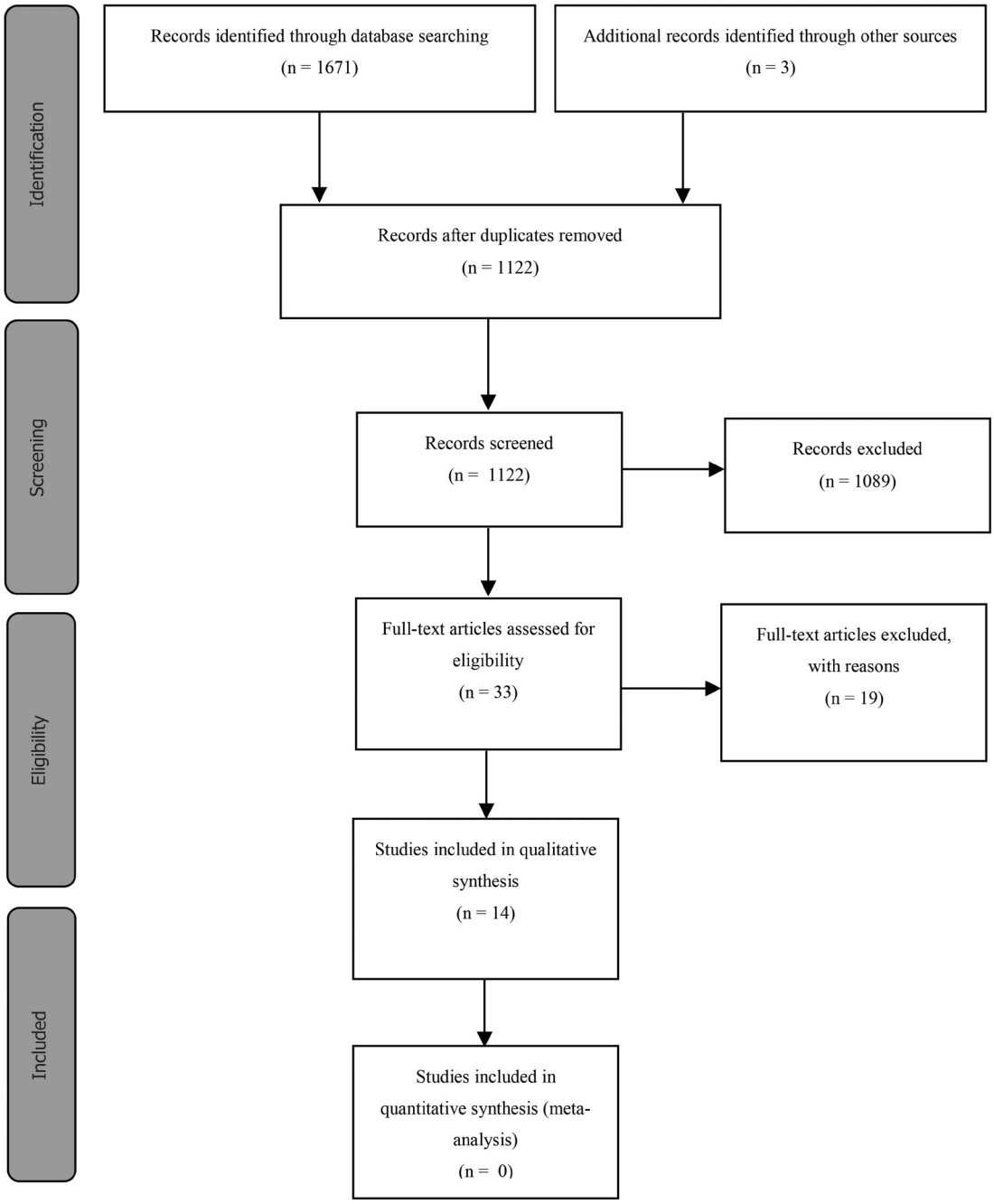


**Figure 1.** Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram.

age was 41.6 years, ranging from 15 years till 64 years old.

### *Intervention characteristics*

The intervention duration was reported in each study, ranging from 1 session to 90 sessions. Most of the studies (n = 7) reported 60 minutes sessions, except for Carew et al. (2007) and Kalra (1978), who described 45 minutes sessions. Some studies (n = 5) did not specify the exact duration in minutes, such as Dacakis (2000), Hancock and Helenius (2012), Hancock and Garabedian (2013), Kawitzky and McAllister (2020) and Mészáros et al. (2005). The content of the intervention sessions included **pitch elevation techniques** (Bralley et al., 1978; Dacakis, 2000; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Kalra, 1978; Kaye et al., 1993; Mészáros et al.,

**Table 3.** Summary of studies reviewed (*results which were marked significant changes in the study).

| Study characteristics | | | | Intervention | | | Outcome | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Article | Study design | Sample size | Age (years) and description of the participant(s) | Intervention duration | Intervention content | Speech task | $f_o$ | Frequency and intensity range | Formant frequencies | Self-perception and satisfaction | Listener ratings | Level of evidence |
| Bralley et al. (1978) | Prospective | 1 | 49<br>Male transsexual | 7 sessions, one hour each | Increasing habitual pitch and pitch range, tasks ranging from production of a single monosyllable to conversational speech. Prevention of vocal abuse in the new mode of production was also of primary concern. | Conversational speech | **Habitual pitch in conversational speech (median)**<br>Pre: 145 Hz<br>Post: 165 Hz*<br>**Fundamental frequency range during conversational speech:**<br>Pre: 30 Hz<br>Post: 70 Hz* | Pitch range:<br>Pre: 100-425 Hz<br>Post: no results | N/A | The client expressed satisfaction with the results attained.<br>More self confidence | 15 judges who were instructed to rate the voices on a l-to-7 scale (1 representing very masculine and 7 representing a very feminine). Control cis female and male samples were included.<br>Pretreatment: 3.7, last session: 4.6* | V |
| Carew et al. (2007) | Prospective | 10 | 40 (25–64)<br>Male-to-female transsexuals | 5 sessions, weekly, 45 minutes each | Targeting oral resonance: lip spreading and increased forward tongue carriage, ranging from isolated vowels to conversation, homework sheets. | The Rainbow Passage | Median pitch during the passage:<br>Pre: 115.2 Hz<br>Post: 129.3 Hz*<br>Mean pitch during the passage:<br>Pre: 119.4 Hz<br>Post: 133.3 Hz* | N/A | Extracted vowels /a/, /i/, and /ʊ/:<br>Higher F1* during /a/ and /ʊ/, F2* during /a/ and F3* for all three vowels | Higher femininity and higher satisfaction*<br>Obtained by VASs | 8 speech pathology student raters, mixed results due to poor interrater reliability. 70% achieved general increases in the perception of femininity (VAS) | V |
| Dacakis (2000) | Retrospective | 10 | 44.5 (32.2–58)<br>Male-to-female transsexuals | 10–90 sessions | Increasing the mean fundamental frequency, no further details. | 2 minutes of spontaneous speech (monologue) | Mean fundamental frequency:<br>Pre: 125.5 Hz*<br>Post: 168.1 Hz*<br>Follow-up: 146.5 Hz*<br>Significant correlation between number of interventions and maintenance of $f_o$ increase. | N/A | N/A | Visual analogue scale (0–100, not at all satisfied – extremely satisfied)<br>Mean results at discharge: 78 (52-100)<br>Mean results at follow up: 80 (49-100) | N/A | V |
| Gelfer and Tice (2013) | Prospective | 5<br>5 cis male and 5 cis female control participants, providing speech samples for analysis and listening experiment | 46.5 (42.5–52.3)<br>Male-to-female (MTF) transgender (TG) clients | 15.4 sessions (8 weeks, two times a week, 60 minutes each) | Group therapy, starting with their target pitch on /m/, /n/, /l/, /r/-initiated consonant-vowel syllables to habituate their new SFF and help facilitate good oral resonance, building up to focusing on pitch, quality, intonation, and pitch range during sentences and spontaneous speech. | Isolated sustained vowels /a/, /i/, /u/, The Rainbow Passage, spontaneous speech, 10 semi spontaneous Q/A sets | **Semi-spontaneous Q/A sets (mean):**<br>Pre: 126 Hz*<br>Post: 210 Hz*<br>Follow-up: 171 Hz*<br>**Spontaneous speech:**<br>Pre: 119 Hz*<br>Post: 178 Hz*<br>Follow-up: 138 Hz*<br>**Rainbow Passage:**<br>Pre: 123 Hz*<br>Post: 194 Hz*<br>Follow-up: 155 Hz* | N/A | **Semi-spontaneous Q/A sets:**<br>F1 /i/<br>Pre: 267 Hz<br>Post: 335 Hz<br>Follow-up: 280 Hz<br>F2 /i/<br>Pre: 2373 Hz<br>Post: 2562 Hz<br>Follow-up: 2561 Hz<br>F3 /i/<br>Pre: 3029 Hz<br>Post: 2977 Hz<br>Follow-up: 3140 Hz<br>Higher F1 during sustained vowel /i/*<br>No significant changes in sustained vowels /a/ and /u/ | N/A | 5 cis male and 5 cis female control participants, 52 college students as listener participants. Female 1.9% of the time in pretest, 50.8% in immediate posttest, 33.1% in long-term posttest (15 months after termination of therapy). Perceived less masculine and more feminine in post tests*. | II |

(Continued)

**Table 3.** (Continued)

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gelfer and Van Dong (2013) | Prospective | 3<br>3 cis male and 3 cis female control participants, providing speech samples for analysis and listening experiment | 43.1 (32.11–50.5)<br>Male-to-female (MTF) transgender (TG) clients | 12 sessions (6 weeks, 60 minutes each) | Individual therapy, targeting pitch by chanting syllables beginning with /m/, /n/, and /l/,…until spontaneous speech (Vocal Function Exercises, VFE), with a home protocol. | The Rainbow Passage, spontaneous speech, 10 semi spontaneous Q/A sets | **Semi-spontaneous Q/A sets (mean):**<br>Pre: 125 Hz<br>Post: 183 Hz<br>**Spontaneous speech:**<br>Pre: 116 Hz<br>Post: 153 Hz<br>**Rainbow Passage:**<br>Pre: 122 Hz<br>Post: 177 Hz<br>(no interferential statistics) | N/A | **Semi-spontaneous Q/A sets:**<br>F1 /i/:<br>Pre: 299 Hz<br>Post: 353 Hz<br>F2 /i/:<br>Pre: 2189 Hz<br>Post: 2323 Hz<br>F3 /i/:<br>Pre: 2640 Hz<br>Post: 2987 Hz<br>(no interferential statistics) | Opinions on the effectiveness of the VFE home protocol, its ease of implementation, overall impressions of therapy. Scale of 1–5. They were generally positive about the addition of VFE to their therapy experience. | 3 cis male and 3 cis female control participants, 27 college students. Identified as male following therapy, although they were rated as less masculine and more feminine*. | II |
| Hancock and Helenius (2012) | Prospective | 1 | 15.3<br>Male-to-female transgender adolescent | 15 sessions (7 months) | Education and counseling concerning vocal hygiene, posture, breathing patterns. Increasing the habitual pitch and using a fuller sounding voice with more forward resonance and normal vocal quality. Slowing speech rate, blending phonemes, and using upward intonation patterns. characterized by consistent, unconstrained vocal tract. Therapy began with relaxation exercises and laryngeal massage followed by the voice goals. | Sustained vowel /a/, The Rainbow Passage, Cookie Theft Picture, spontaneous speech, four corner vowels in /hVd/, frequency range | **Sustained /a/ (mean):**<br>Pre: 205 Hz<br>Post: 209 Hz<br>Follow-up: 209 Hz<br>**Rainbow Passage (mean):**<br>Pre: 158 Hz<br>Post: 204 Hz<br>Follow-up: 204 Hz<br>**Cookie Theft Picture (mean):**<br>Pre: 141 Hz<br>Post: 191 Hz<br>Follow-up: 193 Hz<br>**Spontaneous speech (mean):**<br>Pre: 151 Hz<br>Post: 172 Hz<br>Follow-up: 169 Hz | **Frequency range during spontaneous speech:**<br>Pre: 98–330 Hz<br>Post: 104–349 Hz<br>Follow-up: 69–330 Hz<br>**Total frequency range:**<br>Pre: 87–880 Hz<br>Post: 49–932 Hz<br>Follow-up: 69–392 Hz | Higher F1 and F2 for all vowels*<br>Pre: between male and female<br>Post: toward female values during treatment, except vowel /u/. | TSEQ:<br>Pre: 106/120<br>6 months later: 79/120<br>2 months after therapy ended: 53/120<br>Positive change in attitude and self-perception; an overall satisfaction. | 6 cis females and 4 cis males, VAS from masculine to feminine.<br>Increase from 27 to 79 (increase of 192%) for Cookie Theft Picture, 31 to 72 for spontaneous speech (increase of 132%). | V |
| Hancock and Garabedian (2013) | Retrospective | 25 | 43 (21–60)<br>Male-to-female transgender individuals | 22 sessions (2–77 session, typically once a week) | Increasing fundamental frequency, intonation, resonance, vocabulary, pragmatics, non-verbal communication, reduce phonotraumatic behaviors, vocal hygiene, relaxation techniques and respiration, homework included. | Sustained vowel /a/, The Rainbow Passage, frequency range, spontaneous speech | **Sustained /a/ (mean):**<br>Pre: 136 Hz*<br>Post: 184 Hz*<br>**Rainbow Passage (mean):**<br>Pre: 124 Hz*<br>Post: 156 Hz*<br>**Spontaneous speech (mean):**<br>Pre: 122 Hz*<br>Post: 150 Hz* | **Total frequency range:**<br>Pre: 90–465 Hz<br>Post: 88–579 Hz | N/A | TSEQ: insufficient data, only 4/25 completed the questionnaire at the start and end. Three increased, one remained stable | N/A | V |

(Continued)

**Table 3.** (Continued)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kalra (1978) | Prospective | 1 | 27<br>Morphological male transsexual who became a female | 3 months, weekly session of 45 minutes | Pitch elevation: exercises which reinforced resonance and maintained a balance between the vocal generator and supraglottal resonators. As new pitch levels were acquired, Froeschel's chewing method was used to increase anterior oral resonance. Reducing clavicular breathing patterns and lessen pharyngeal tension. | Spontaneous speech, automatic speech tasks, chewing | **Chewing (mean):**<br>Pre: 180 Hz<br>Post: 200 Hz<br>Follow-up: 190 Hz<br>**Days of the week while chewing (mean):**<br>Pre: 180 Hz<br>Post: 196 Hz<br>Follow-up: 185 Hz<br>**Months of the year while chewing (mean):**<br>Pre: 160 Hz<br>Post: 190 Hz<br>Follow-up: 185 Hz<br>**Spontaneous speech (mean):**<br>Pre: 153 Hz<br>Post: 198 Hz<br>Follow-up: 200 Hz<br>No statistical data | N/A | N/A | N/A | N/A | V |
| Kawitzky and McAllister (2020) | Prospective | 12<br>19 cis men as control participants | 45.2 (21–71)<br>Transgender women | 1 session | Increasing F2 through biofeedback. Vowels /ae/, /ʌ/, and /ɑ/ were target vowels, into the context of real words, surrounded by voiced plosives. Three conditions: unshifted, shifted-up and shifted down position. Participants were informed of two strategies they might use to manipulate the location ofthe peak: (a) changing the positioning of the tongue in the mouth, and (b) changing the shape of the lips. | Vowels /ae/, /ʌ/, and /ɑ/, target words bud/bad/bod | Fundamental frequency during target words: visual inspection of boxplots does not indicate any reliable patterns of $f_o$ differences across conditions. | N/A | **F1:**<br>Some differences were observed between shifted-up, shifted-down, and unshifted target conditions, but were generally small in magnitude and showed no clear pattern across vowel targets. Shifted-down was lower dan unshifted*, but shifted-up was not higher than unshifted.<br>**F2:**<br>Differed consistently across elicitation conditions. Higher F2 during shifted-up*, lower F2 during shifted-down*.<br>**F3:**<br>No differences between conditions. | N/A | 26 blinded naïve raters on VAS (definitely male to definitely female). Higher F2 or $f_o$ values received higher perceptual ratings of femininity. Best to combine both. Combining F1 or F3 with $f_o$ yielded higher femininity ratings* but not on their own. | II |

(Continued)

**Table 3.** (Continued)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kaye et al. (1993) | Prospective | 1 | 44 Male-to-female transsexual subject | 12 sessions, 60 minutes each | Raising fundamental frequency, while ensuring that the procedures would not cause laryngeal strain. Correct sound production and resonance were also emphasized. | First two sentences of The Rainbow Passage | **Mean during reading:** Pre: 101 Hz Post: 135 Hz **Frequency range during reading:** Pre: 21 Hz Post: 31 Hz | N/A | N/A | N/A | 2 cis male, 2 cis female, 1 trans male control participants, 40 student listeners. Two 5-point equal interval rating scales: femininity and masculinity. Perceived as less masculine* and more feminine* after the therapy, even though never labeled as a female. | V |
| Mészáros et al. (2005) | Prospective | 3 | 23 (20–26) Male-to-Female transsexuals | 9.6 months (8–19 months, one session a week) | Developing a female pitch, high-pitch vocalization by using the first articulation zone, a costo-abdominal breathing technique, decreasing laryngeal muscle tension, and to increase head resonance as compared to chest resonance. Attention was paid to the formation of soft tones, precise articulation and the attainment of a female intonation. | The North Wind and The Sun, voice range profile | Habitual pitch during the North Wind and The Sun (mean of three participants): Pre: 150.6 Hz Post: 191.3 Hz No statistical data | Pitch range diminished on average, as evidenced by an elevation in the lower limit of the pitch range. The vocal intensity range reached normal values. | N/A | The degree to which the person's voice could be used for communication, a subjective scale, 0 = no limitations, 1 = limited communication only in the case of voice loading, 2 = a moderate degree of constant limitation, and 3 = constant severe limitation in everyday communication. All participants went from 3 to 0. | N/A | V |
| Mount and Salmon (1988) | Prospective | 1 | 63 Male-to-female transsexual | 88 sessions (60 minutes each, over 11 months period) | Higher pitch levels while avoiding vocal abuse, and modify tongue carriage to achieve higher resonance characteristics of the vocal tract. Additional: promote a breathy vocal attack, and establish appropriate inflection patterns at higher pitch levels. Homework included. | Conversational speech, sustained vowels, audio and video recorded. | **Sustained vowel /i/ (mean):** Pre: 110 Hz Post: 195 Hz Follow-up: 235 Hz **Sustained vowel /a/ (mean):** Pre: 110 Hz Post: 210 Hz Follow-up: 230 Hz **Sustained vowel /u/ (mean):** Pre: 110 Hz Post: 210 Hz Follow-up: 200 Hz No statistical data | N/A | **Pre – post - follow-up (after 5 years): F1 during /i/:** 222-278-371 Hz **F1 during /a/:** 606-829-955 Hz **F1 during /u/:** 277-322-389 Hz **F2 during /i/:** 2092-2383-2389 Hz **F2 during /a/:** 1053-1147-1191 Hz **F2 during /u/:** 810-1065-1168 Hz **F3 during /i/:** 3467-3505-3363 Hz **F3 during /a/:** 2331-2718-2853 Hz **F3 during /u/:** 1944-2458-2151 Hz No statistical data | N/A | N/A | V |

(Continued)

**Table 3.** (Continued)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Quinn and Swain (2018) | Prospective | 1 | 17 Transgender individual | 2 60 minutes sessions a day, for two weeks | Vocal health and flexibility, pitch raising through Stemple's Vocal Function Exercises, oral resonance target through the Katherine Verdolini's Lessac-Madsen Resonant Voice Therapy program (focus on lip spreading and anterior tongue carriage). Additional: vocal education (gentle onsets), feminine intonation patterns. Homework included. | The Rainbow Passage, The Grandfather Passage, spontaneous speech, frequency range during reading | **Rainbow Passage (mean):** Pre: 145 Hz Post: 179 Hz **Grandfather Passage (mean):** Pre: 146 Hz Post: 166 Hz **Spontaneous speech (mean):** Pre: 143 Hz Post: 144 Hz No statistical data | **Rainbow Passage:** Pre: 73 Hz Post: 80 Hz **Grandfather Passage:** Pre: 73 Hz Post: 81 Hz **Spontaneous speech:** Pre: 75 Hz Post: 78 Hz No statistical data | N/A | **Self-rated femininity and satisfaction** (VAS of 10 cm): increase during voice intervention phase and after ceasing intervention (reading and spontaneous speech)*. **TVQ:** Pre: 100/120 Post: 81/120 **Short interview** concerning their experiences completing the program: 9/10 on the usefulness of the therapy program, and 7/10 to what degree of change they perceived in their voice. | 7 external raters evaluated femininity (VAS of 10 cm): increase in femininity during voice intervention phase*. Oral resonance was found to have no additional statistically significant impact on outcomes. More feminine in reading than spontaneous speech. No statistically significant difference between listener and self-ratings*. | V |
| Söderpalm et al. (2004) | Retrospective | 21 | 40 Male-to-female transsexual individuals | 17 sessions (3–45), 45–60 minutes each | Vocal hygiene exercises, pitch raising exercises by gradual climbing, 'gliding', precise and more anterior articulation. | A Difficult Case, sustained vowels, interview | Less than half the patients produced a fundamental frequency above 155 Hz at the end of therapy but ten patients had reached this level at follow-up. There was an increase of 20 Hz for 10 patients between baseline and follow-up measures of fundamental frequency*. Increase in mean fundamental frequency is greater for those with more than 14 sessions. | N/A | N/A | VAS of 10 cm (masculinity/femininity): Pre: 6.6 Follow-up: 8 | N/A | V |

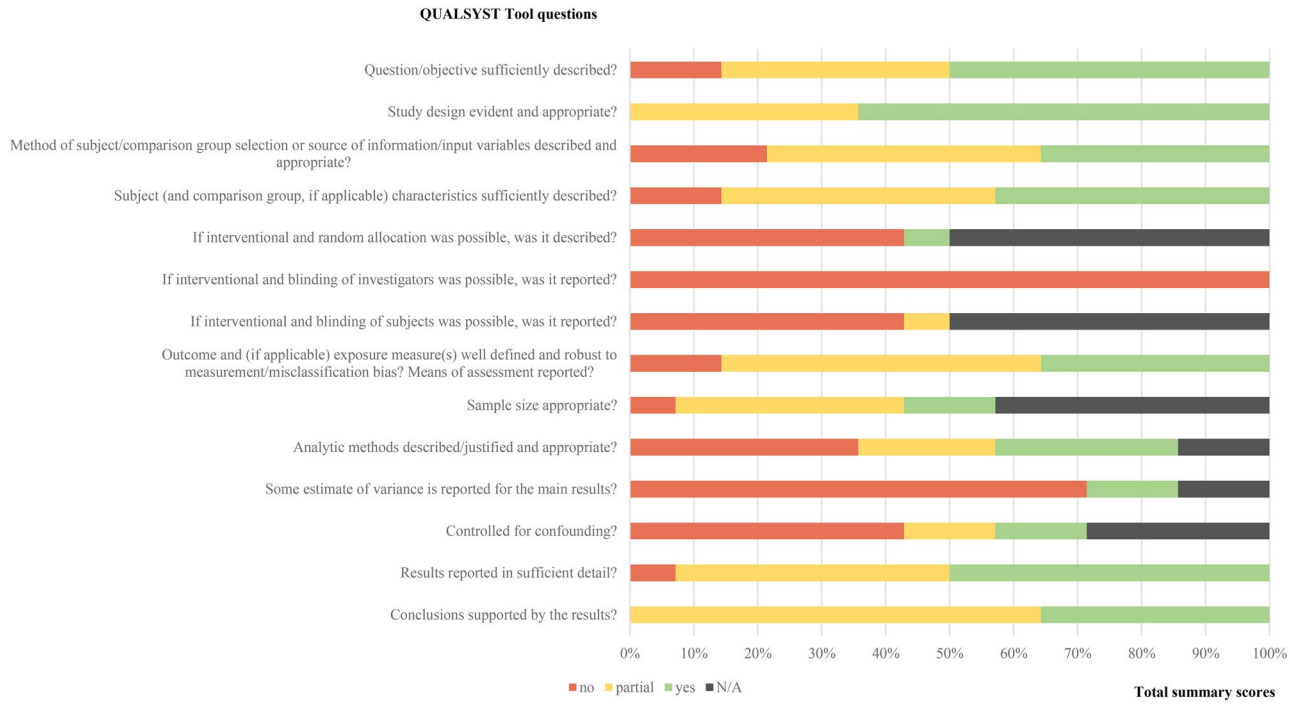


**Figure 2.** Risk of bias of the individual studies: results of the QUALSYST Tool.

2005; Mount & Salmon, 1988; Quinn & Swain, 2018; Söderpalm et al., 2004), **oral resonance training** (Carew et al., 2007; Gelfer & Tice, 2013; Hancock & Garabedian, 2013; Kalra, 1978; Kawitzky & McAllister, 2020; Mészáros et al., 2005; Mount & Salmon, 1988; Quinn & Swain, 2018; Söderpalm et al., 2004), **prosody** (Gelfer & Tice, 2013; Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Mészáros et al., 2005; Mount & Salmon, 1988; Quinn & Swain, 2018), **vocal quality** (Gelfer & Tice, 2013; Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Kaye et al., 1993; Mészáros et al., 2005; Mount & Salmon, 1988), **vocal hygiene** (Bralley et al., 1978; Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Mount & Salmon, 1988; Quinn & Swain, 2018; Söderpalm et al., 2004), **breathing patterns** (Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Kalra, 1978; Mészáros et al., 2005) and **non-verbal communication** (Hancock & Garabedian, 2013).

Five of the fourteen studies reported motivating their participants to practice at home with some **carry-over tasks** (Carew et al., 2007; Gelfer & Van Dong, 2013; Hancock & Garabedian, 2013; Mount & Salmon, 1988; Quinn & Swain, 2018). All of the studies choose individual therapy, except for Gelfer and Tice (2013) who reported using a group setting for their therapy protocol.

Gelfer and Van Dong (2013) and Quinn and Swain (2018) implemented an existing therapy protocol, the Vocal Function Exercises (Stemple, 1984). Kalra (1978) mentioned Froeschel's chewing method (Froeschels, 1952) in their protocol and Quinn and Swain (2018) used the Resonant voice therapy program by Katherine Verdolini (Verdolini Abbott, 2008).

## *Outcome characteristics*

### *Fundamental frequency ($f_o$)*

Fundamental frequency ($f_o$) during sustained vowel. All studies who reported the $f_o$ during sustained vowel(s) (n = 3) described an increase of $f_o$, varying between 4 Hz and 100 Hz, or between 0.34 and 11 semitones (ST), with a mean of 7 ST. Hancock and Helenius (2012), a case study, reported an increase of 4 Hz during sustained vowel /a:/ (205 Hz pre – 209 Hz post – 209 Hz follow-up, pre – post: 0.34 ST, pre – follow-up: 0.34 ST). During the retrospective study by Hancock and Garabedian (2013) on the other hand, an increase of 48 Hz (136 Hz pre

– 184 Hz post, 5 ST) was reported during sustained vowel /a/. The case study by Mount and Salmon (1988) described both vowels /a/, /i/ and /u/, with an increasing $f_o$ of respectively 110 Hz (pre) – 210 Hz (post) – 230 Hz (follow-up) (pre – post: 11 ST, pre – follow-up: 13 ST), 110 Hz (pre) – 195 Hz (post) – 235 Hz (follow-up) (pre – post: 10 ST, pre – follow-up: 13 ST) and 110 Hz (pre) – 210 Hz (post) – 200 Hz (follow-up) (pre – post: 11 ST, pre – follow-up: 10 ST).

***Fundamental frequency ($f_o$) during reading.*** Of the 10 studies reporting $f_o$ during reading, 8 used the Rainbow Passage (Fairbanks, 1960) to investigate the $f_o$ during reading. A mean increase of $f_o$ was observed of 5 ST. Carew et al. (2007) reported an increase of 14.1 Hz, going from 115.2 to 129.3 Hz (2 ST). Gelfer and Tice (2013) described an increase of 71 Hz after the post measurement (123 Hz pre – 194 Hz post, 8 ST), but then again a drop in pitch at follow-up (155 Hz, pre – follow-up: 4 ST). Gelfer and Van Dong (2013) did not report a follow-up measurement and had a 55 Hz increase of pitch (122 Hz pre – 177 Hz post, 6 ST). Hancock and Helenius (2012) increased the $f_o$ with 46 Hz during their case study, (158 Hz pre – 204 Hz post, 4 ST) and stayed stable at follow-up (204 Hz, 4 ST). The other study by Hancock and Garabedian (2013) reported an increase of 32 Hz (124 Hz pre – 156 Hz post, 4 ST). Kawitzky and McAllister (2020) did not focus on raising the pitch in their intervention, and did not find any reliable patterns in $f_o$ differences. Kaye et al. (1993) had a change of pitch of 34 Hz (n = 1) in total (101 Hz pre – 135 Hz post, 4 ST). Mészáros et al. (2005) used another phonetically balanced text, the North wind and the Sun, and reported an increase of 40.7 Hz (150.6 Hz pre – 191.3 Hz post, 4 ST). Quinn and Swain (2018) described an increase of pitch of 34 Hz of their single case (145 Hz pre – 179 Hz post, 4 ST) during the Rainbow Passage. Lastly, Söderpalm et al. (2004) noticed an increase 20 Hz during reading and a greater increase for those with more than 14 sessions.

***Fundamental frequency ($f_o$) during spontaneous/conversational speech.*** 8 studies investigated the $f_o$ during spontaneous or conversational speech. There was a mean general increase of 4 ST pre – post treatment. Bralley et al. (1978) reported a 20 Hz increase during conversation speech of their participant, going from 145 Hz to 165 Hz (2 ST), whereas Dacakis (2000) showed an increase of 42.6 Hz at the post measurement, 125.5 Hz to 168.1 Hz (5 ST), but decreasing again at follow-up until 146.5 Hz (pre – follow-up: 3 ST). The authors mentioned a correlation between the number of interventions and the maintenance of the $f_o$ increase, i.e. a higher number of interventions caused a higher $f_o$ increase. Both Gelfer and Tice (2013) and Gelfer and Van Dong (2013) mentioned an increase, respectively 59 Hz (119 Hz pre to 178 Hz post, 7 ST) and 37 Hz (116 Hz pre to 153 Hz post, 5 ST), with a drawback during follow-up to 19 Hz (138 Hz, pre – follow-up: 3 ST) in the study by Gelfer and Tice (2013). Hancock and Helenius (2012) on the other hand described a stable increase in $f_o$ during follow-up of their case study (151 to 172 to 169 Hz, pre – post: 2 ST, pre – follow-up: 2 ST). An increase of 28 Hz (122 Hz pre to 150 Hz post, 4 ST) was reported in the study by Hancock and Garabedian (2013) and 45 Hz (153 Hz pre and 198 Hz post, with 200 Hz during follow-up, pre – post: 4 ST, pre – follow-up: 5 ST) in the case study by Kalra (1978). Quinn and Swain (2018) did not see a high increase during spontaneous speech of the participant, 1 Hz (143 Hz pre to 144 Hz post, 0.12 ST).

### *Total frequency and intensity range*

Not all studies described frequency and intensity ranges. Of the 4 studies reporting total frequency and intensity range, 2 mentioned pre and post values. Hancock and Helenius (2012) reported a total frequency range of their case study of 87-880 Hz (pre, 40 ST) to 49-932 Hz (post, 51 ST) and 69-392 Hz (follow-up, 30 ST) (pre – post: 11 ST, pre – follow-up: –10 ST). The retrospective study by Hancock and Garabedian (2013) of 25 participants showed the total frequency range increasing, from 90-465 Hz (pre, 28 ST) to 88-579 Hz (post, 33 ST) (pre – post: 5 ST). Bralley et al. (1978) on the other hand reported a total frequency range from 100 Hz till 425 Hz (25 ST) during the pre training measurement of their subject. Mészáros et al. (2005) mentioned a diminishing pitch range, caused by an elevation of the lower limit.

### *Frequency and intensity range during continuous speech*

3 studies investigated the frequency and intensity range during continuous speech. Bralley et al. (1978) reported a frequency range during conversational speech of 30 Hz (pre) and 70 Hz (post) (15 ST). Hancock and Helenius (2012) showed the frequency range during spontaneous speech changing from 98-330 Hz (pre, 21 ST) to 104-349 Hz (post, 21 ST) and 69-330 Hz (follow-up, 27 ST) (pre – post: 0 ST, pre – follow-up: 6 ST). Quinn and Swain (2018) reported slightly increased pitch ranges during conversational and spontaneous speech. It is important to acknowledge that all three of these studies concern single cases.

### *Formant frequencies*

Only 6 out of the 14 included studies investigated the formant frequencies before and after the intervention. Carew et al. (2007) extracted vowels /a/, /i/, and /ʊ/ from the Rainbow Passage and Gelfer and Tice (2013) and Gelfer and Van Dong (2013) extracted vowel /i/ from their semi spontaneous Q/A sets and investigated the formant frequencies in sustained vowels /a/, /i/ and /u/. The other studies used solely sustained vowels (Hancock & Helenius, 2012; Mount & Salmon, 1988) or target words (Kawitzky and McAllister, 2020). Higher F1 values during /a/ and /ʊ/ (Carew et al., 2007), /i/ (Gelfer & Tice, 2013), all vowels (Hancock & Helenius, 2012), higher F2 values during /a/ (Carew et al., 2007), all vowels (Hancock & Helenius, 2012) and higher F3 for all three vowels (Carew et al., 2007) were found. Kawitzky and McAllister (2020) mentioned that participants were able to significantly lower their F1 and increase and decrease their F2. Some other studies reported formant frequencies, but these were not analyzed statistically.

### *Self-perception and satisfaction*

10 of the 14 studies reported data concerning the self-perception and satisfaction of the participants. These data were obtained through an interview (Bralley et al., 1978; Quinn & Swain, 2018), visual analogue scales (Carew et al., 2007; Dacakis, 2000; Quinn & Swain, 2018; Söderpalm et al., 2004), Likert scales (Gelfer & Van Dong, 2013; Mészáros et al., 2005) and standardized questionnaires such as the Transsexual Self-Evaluation Questionnaire (TSEQ) by Davies and Goldberg (2006) (Hancock & Garabedian, 2013; Hancock & Helenius, 2012) and the Transsexual Voice Questionnaire (TVQ) by Dacakis et al. (2013) (Quinn & Swain, 2018). The studies reported increased satisfaction (Bralley et al., 1978; Carew et al., 2007; Dacakis, 2000; Gelfer & Van Dong, 2013; Hancock & Helenius, 2012; Quinn & Swain, 2018), more self-confidence (Bralley et al., 1978; Hancock & Helenius, 2012), higher self-perception of femininity (Carew et al., 2007; Quinn & Swain, 2018; Söderpalm et al., 2004), lower scores on the TSEQ, TVQ or Likert scale, i.e. lower impact on the psychosocial functioning or limitations concerning communication (Hancock & Helenius, 2012; Mészáros et al., 2005; Quinn & Swain, 2018).

### *Listener ratings*

Several studies described significant increases in femininity and decreases in masculinity on visual analogue scales, collected through listening experiments (Bralley et al., 1978; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Hancock & Helenius, 2012; Kaye et al., 1993; Quinn & Swain, 2018). Due to poor interrater reliability, mixed results were reported by Carew et al. (2007). Binary gender identification results revealed a higher number of female identification post treatment and follow-up in the study by Gelfer and Tice (2013). In the study by Gelfer and Van Dong (2013) and Kaye et al. (1993), participants were still rated as male following therapy. Kawitzky and McAllister (2020) concluded that higher F2 or $f_o$ values received higher perceptual ratings of femininity. Combining F1 or F3 with $f_o$ yielded higher femininity ratings but not on their own.

## Discussion

This systematic review was performed in order to provide speech and language pathologists with the current literature concerning the effects of speech therapy in transgender women in terms of acoustic and perceptual outcomes. It is important to give speech language pathologists an

overview of the existing literature concerning the effects of speech therapy in transgender women in order to establish a voice and communication which is congruent with their gender identity. Previous reviews concerning the effectiveness of speech therapy for transgender women mentioned the importance of communication characteristics that contribute to perceptions of gender, such as pitch, resonance and articulation (Dacakis et al., 2012; Davies et al., 2015; Oates & Dacakis, 1983). Additionally, they reported several recommendations for speech language pathologists working in this field. However, this is the first systematic review that has summarized the evidence of the effects of speech therapy, using a strict systematic methodology of the PRISMA guidelines (Moher et al., 2015) and including recent literature. 14 studies have been identified during the search in this systematic review, of which 6 were case studies and 3 had retrospective designs. There were a total of 95 participants, ranging from 15 to 64 years old (mean 41.6 years). Due to the number of participants and study designs that were reported, it can be concluded that research on speech therapy outcomes for transgender women is quite limited. Results are somewhat promising and suggest that speech therapy could result in vocal changes and gender perception.

Looking at the content of the intervention, a lot of attention has been paid to several aspects which contribute to gender perception of speaker gender, described by Leung et al. (2018), such as pitch and resonance (Bralley et al., 1978; Carew et al., 2007; Dacakis, 2000; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Kalra, 1978; Kawitzky & McAllister, 2020; Kaye et al., 1993; Mészáros et al., 2005; Mount & Salmon, 1988; Quinn & Swain, 2018; Söderpalm et al., 2004). It is noticeable that therapy goals seem to change over the years, i.e. the older the study, the more focus is laid on pitch elevation, in comparison to more recent literature, who regularly include the aspects of resonance. However, another aspect which is associated with gender perception of speaker gender is intonation (Leung et al., 2018). Only six studies addressed intonation in their intervention content (Gelfer & Tice, 2013; Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Mészáros et al., 2005; Mount & Salmon, 1988; Quinn & Swain, 2018). Overall, most of the studies reported a combination of different therapy goals, concluding in a difficult interpretation of the results.

The included studies in this review reported various durations of the intervention, ranging from 1 session to 90 sessions. Seven studies described 60 minutes sessions, except for Carew et al. (2007) and Kalra (1978), who described 45 minutes sessions. Some studies (n = 5) did not specify the exact duration in minutes, such as Dacakis (2000), Hancock and Helenius (2012), Hancock and Garabedian (2013), Kawitzky and McAllister (2020) and Mészáros et al. (2005). De Bodt et al. (2015) found that published voice therapy described 10.87 sessions of mostly 30 or 60 minutes. Söderpalm et al. (2004) addressed the fact that participants with more than 14 intervention sessions had a greater increase in $f_o$ immediately after the intervention and at the follow-up appointment (Dacakis, 2000; Gelfer & Tice, 2013). In the study by Meerschman et al. (2019), short-term intensive voice therapy is at least equally effective in treating patients with dysphonia as long-term traditional voice therapy. The intensive program made an equal progress in only 2 weeks and 12 hours of therapy compared with the traditional long-term program that needed 6 months and 24 hours of therapy. Although these studies included a study population of patients with dysphonia, it is important to discuss the intervention duration of voice training for transgender women. Cost-effectiveness and session attendance might be potential advantages of an intensive intervention program for this population. Future research should compare these intervention characteristics.

Concerning outcome characteristics, fundamental frequency ($f_o$) during sustained vowels, reading and spontaneous speech is the most described parameter in this systematic review. Most of the studies reported mean values of the $f_o$ (n = 11), one study reported median values (Bralley et al., 1978) and one study reported both mean and median values (Carew et al., 2007). Mean values are more sensitive to the existence of outliers than the median (Leys et al., 2013). The results are hard to interpret as they include a scattered

range of $f_o$ increases. Sustained vowel /a/ for example increased 4 Hz in the study by Hancock and Helenius (2012), to 48 Hz in the study by Hancock and Garabedian (2013), to 110 Hz in the study by Mount and Salmon (1988), with a mean increase of 7 ST. It is known that producing a sustained vowel might vary between different elicitations. Vocal fluctuations related to voice onset, voice termination and voice breaks, can have a relatively large influence on short signals (Maryn et al., 2010). It is important to give strict instructions across these elicitations, such as "at comfortable loudness and pitch" (Dejonckere et al., 2001; Fitch, 1990). The mean $f_o$ during reading was described with increases ranging from 14 to 71 Hz immediately after the intervention, with most studies reporting an increase between 30 and 45 Hz, with a mean of 5 ST (Gelfer & Van Dong, 2013; Hancock & Garabedian, 2013; Hancock & Helenius, 2012; Kaye et al., 1993; Mészáros et al., 2005; Quinn & Swain, 2018). Studies who described longer-term follow-up (between 2 months and 8;9 years after discharge) reported a drop of $f_o$ during the follow-up measurement (Gelfer & Tice, 2013), revealing a need for more generalization of the increased pitch during their speech. During spontaneous or conversational speech, increases have been reported from 1 to 59 Hz, with a mean of 4 ST. Quinn and Swain (2018), who reported the 1 Hz difference, stated that "in comparison to a reading task, spontaneous speech is associated with additional cognitive load and puts additional strain on a speaker's ability to self-regulate their behavior. Depleted self-regulation has been shown to negatively impact performance in voice tasks that require active behavioral modification". As their participant showed impaired executive functioning, the minimal increase might be explained by this. Similar to the results during reading, a set-back has been reported at follow-up measurements (Dacakis, 2000; Gelfer & Tice, 2013).

It is important to acknowledge that in general, most $f_o$ post measurements are still in the gender ambiguous zone (150 Hz – 185 Hz, Mordaunt (2006)). Not only raising the speaking pitch to a value higher than 180 Hz is necessary in order to be perceived female during gender perception. The findings in the systematic review by Leung et al. (2018) suggested that speaking in the range of 140 Hz as a lower limit and 300 Hz as an upper limit would also contribute to listener perceptions that the speaker is female. However, not a lot of studies in this review reported frequency range characteristics. Bralley et al. (1978) described a frequency range during conversational speech (30 Hz pre – 70 Hz post) but did not describe the upper and lower limits. Hancock and Helenius (2012) did not reach the lower limit of 140 Hz (104 Hz post) but reported a higher upper limit than 300 Hz (349 Hz post). Reporting these values in future research might be beneficial for explaining the contribution of $f_o$ in gender perception of speaker gender.

Resonance of the vocal tract can be described by formant frequencies and are determined by the length and shape of the vocal tract (Shriberg & Kent, 2003). Formant frequencies represent an objective measure that may be useful in studying the effects of treatment on vocal function (Kayikci et al., 2012). They are the resonant harmonics in the speech spectrum and are described as being the characteristic partials that help identify the vowel to the listener (Atal & Hanauer, 1971; Baken & Orlikoff, 2000). As has been shown in the review by Leung et al. (2018), the frequency of the first 4 formants have been empirically shown to contribute to gender perception. Looking at the results of the formant frequencies, most studies used the frequencies extracted from sustained vowels. Only Carew et al. (2007), Gelfer and Tice (2013) and Gelfer and Van Dong (2013) extracted vowels from reading or spontaneous speech. When extracting vowels from continuous speech, they are more representative for actual daily communication. It should be mentioned that a well described protocol of extracting and analyzing formant frequencies is essential for reproducibility of the study. For the first three formant frequencies, increases have been observed, but not for all vowels and not each formant in each study. The study by Carew et al. (2007) targeted forward tongue carriage and lip spreading and observed a general increase in F1, F2, and F3 for all vowels (/i/, /a/, and /ʊ/), but these were only statistically significant for F1 values of /a/ and /ʊ/, F2 values of /a/, and F3 values for all three vowels. Vowel /a/ is the most backed

vowel in Australian English, which might explain the significantly increased F2 of /a/. The authors mentioned that the technique of lip spreading, correlated with F3, appeared to be easily comprehended by their clients in five therapy sessions and that the clients lowered their habitual tongue height during the forward tongue carriage exercises, resulting in the overall increase in F1 values. In the study by Gelfer and Tice (2013) F1 of vowel /i/ (F1) varied significantly among all measurements. These results revealed that participants increased F1 of /i/ in the immediate posttest, decreased significantly in the long-term posttest, but ended up significantly higher than that they had started in the pretest. Hancock and Helenius (2012) showed significantly increased F1 and F2 values for all vowels, caused by the participant's use of more forward articulatory placement to achieve a head resonance rather than chest resonance. Despite the limited nature of the training and practice they received, participants in the study by Kawitzky and McAllister (2020) were generally successful in shifting their F2 frequencies in the direction of a target with visual feedback. Mount and Salmon (1988) also targeted the forward tongue position and reported that it took 11 months before a significant change in F2 values was established. Consequently, it might be possible that resonance outcomes can be altered to support a more feminine perception of the voice.

Psychosocial functioning can be negatively impacted when communication and speech characteristics of transgender people are not congruent with their gender expressions (Colton & Casper, 1996). How others perceive one's gender and femininity is also related to one's self-perceptions of femininity and happiness. McNeill et al. (2008) and Hancock et al. (2011) found that a listener's perception of femininity was positively correlated to the client's perception of her own femininity. All studies in this review who investigated the self-perception and satisfaction of the participant, discovered a higher feminine self-perception of the voice and higher satisfaction after the intervention. Hancock and Garabedian (2013) reported having no sufficient data concerning the TSEQ, which might be associated with their retrospective design. Not all participants reached their goal of 'a very feminine voice' (Quinn & Swain, 2018), but still made improvements in self-perception from some negative feelings they experienced pretreatment. Vocal satisfaction and self-perception might be a feasible treatment goal in itself, independent of a client's goals around passing and socializing (Quinn & Swain, 2018). Intervention that targets the client's self-perception in addition to focusing on the acoustic aspects is recommended for the development of a feminine voice (Hancock & Helenius, 2012).

Six out of eight studies who conducted a listening experiment observed increases in femininity and decreases in masculinity during listening experiments. These results are mainly collected with Likert scales and visual analogue scales. Binary gender identification has been a way to investigate the listener perception of speaker gender as well, with a higher number of female identification post treatment and follow-up in the study by Gelfer and Tice (2013). However, 92.6% of the speech samples in the study by Gelfer and Van Dong (2013) were still rated male post treatment and during the post test of the case study by Kaye et al. (1993) the participant was never labeled as female. They both included 12 intervention sessions, but with very small sample sizes (resp. $n = 3$ and $n = 1$). Four studies incorporated control speech samples in their listening experiment (cisgender male and female speakers) to distract the listeners from the objective of the study in order to avoid biased answers as much as possible (Bralley et al., 1978; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Kaye et al., 1993). The number of listener participants ranged from 8 to 52. Bralley et al. (1978) did not report any information concerning their 15 judges. Carew et al. (2007) included 8 speech pathology students as listeners but observed poor inter-rater reliability results and therefore suggested to use speech pathologists who are experienced in the area of voice to carry out perceptual ratings. Gelfer and Tice (2013) and Gelfer and Van Dong (2013) on the other hand used respectively 52 and 27 college students with no experience in speech pathology. Hancock and Helenius (2012), Kawitzky and McAllister (2020) and Quinn and Swain (2018) only included listeners who met

criteria for intra-rater reliability, respectively 10 college students, 26 blind naïve raters and 7 speech pathology and non-speech pathology students. Forty naïve student listeners were reported by Kaye et al. (1993).

Oates (2006) previously stated that the evidence for effectiveness of voice therapy for transgender clients is weak, with 83% of the studies being "at the very lowest level on the evidence hierarchy. The remaining 17% of publications in this field provide only marginally stronger evidence". Therefore, during the process of writing a systematic review, it is important to assess the quality of the included studies. Both risk of bias and levels of evidence were investigated for this systematic review. The risk of bias assessment was performed with the QUALSYST tool (Kmet et al., 2004). As there is no standard, empirically grounded quality assessment tool suitable for use with a variety of study designs, the authors developed a scoring system to assess the quality of quantitative research reports. Fourteen items were evaluated: sufficiently described objective, appropriate study design, well reported subject selection and subject characteristics, random allocation, blinded investigators and subjects, well defined outcome measures, appropriate sample size, justified analytic methods, reported estimate of variance, controlled for confounding, sufficiently detailed results and conclusions which are supported by the results. The tool revealed total summary scores ranging from 18% to 82%, with a mean score of 49%. When looking at the different questions of the tool, all studies reported a 'no' on the question concerning the blinding of the investigators. Risk for investigator bias is possible and should therefore be avoided by using investigators blinded to the study process (Meerschman et al., 2019). The poor quality of these studies makes it hard to interpret the acoustic and perceptual effects of the interventions, caused by methodological issues. Firstly, three studies have retrospective study designs (Dacakis, 2000; Hancock & Garabedian, 2013; Söderpalm et al., 2004). Secondly, small samples sizes (varying between n = 1 and n = 25) have been reported for each study. In fact, 6 of the 14 studies were case studies (Bralley et al., 1978; Hancock & Helenius, 2012; Kalra, 1978; Kaye et al., 1993; Mount & Salmon, 1988; Quinn & Swain, 2018). Thirdly, some of the studies showed vaguely described therapy contents such as the retrospective studies by Dacakis (2000) and Söderpalm et al. (2004). Dacakis (2000) for example explained the therapy procedures as "the therapeutic techniques employed with all participants were similar, and focused primarily on increasing mean fundamental frequency". This reduces the possibility to reproduce the study and to obtain clear therapy goals to be used by speech language pathologists. Fourthly, long-term follow-up measurements are often absent, which negatively impacts the opportunity to investigate long-term effects of the intervention. Therefore the study misses the chance to check whether further guidance is needed by the speech language pathologist. A last aspect which is a clear limitation in these studies, is the risk for experimenter bias. Furthermore, the levels of evidence of the individual studies were evaluated as well, using the scoring system by Sackett (1989). As most (n = 11) of the studies did not include control participants during their intervention or analysis, a level V was assigned. However, some studies (n = 4) included control participants during their listening experiment. These findings show that there is a definite need for randomized controlled trials.

When transgender women perceive their voice to be incongruent with their gender role, they either go to a speech therapist, undergo phonosurgery or choose to have no intervention (Nolan et al., 2019). It is very important as a client, clinician, speech therapist, ENT doctor or any healthcare provider to know the effects of speech therapy. A difficulty in investigating the effectiveness of any kind of intervention in transgender persons is that therapy goals are often individually determined as the needs of these persons are often different. However, a good clinical trial must contain well-described methodology to increase reproducibility. Nevertheless, there is a clear need for more research that shows whether speech therapy is effective, which exercise contributes to which aspect, and what the ideal duration of an intervention is.

## Conclusion

Research on speech therapy outcomes for transgender women is limited. Preliminary results of pilot studies in transgender women are promising and suggest that speech therapy could result in vocal changes and gender perception. The results of the included studies are sometimes difficult to interpret and compare due to methodological issues. There is an urgent need for effectiveness studies using randomized controlled study designs, larger sample sizes, multidimensional voice assessments including both objective, perceptual and self-rating outcomes, complete and well-described therapy programs, investigators blinded to study process, and longer-term follow-up data.

## Disclosure statement

Conflict of Interest: the authors declare that they have no conflict of interest.

## ORCID

Peter Tomassen http://orcid.org/0000-0002-0698-6574

## References

Atal, B. S., & Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, *50*(2B), 637–655.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.

Bralley, R. C., Bull, G. L., Gore, C. H., & Edgerton, M. T. (1978). Evaluation of vocal pitch in male transsexuals. *Journal of Communication Disorders*, *11*(5), 443–449. https://doi.org/10.1016/0021-9924(78)90037-0

Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery*, *128*(1), 305–310. https://doi.org/10.1097/PRS.0b013e318219c171

Carew, L., Dacakis, G., & Oates, J. (2007). The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *Journal of Voice*, *21*(5), 591–603. https://doi.org/10.1016/j.jvoice.2006.05.005

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PloS One*, *7*(2), e31353.

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, *14*(3), 565–577. https://doi.org/10.1044/jshr.1403.565

Coleman, R. O. (1983). Acoustic correlates of speaker sex identification: Implications for the transsexual voice. *The Journal of Sex Research*, *19*(3), 293–295. https://doi.org/10.1080/00224498309551189

Colton, R., & Casper, J. (1996). Understanding voice problems: A physiological perspective for diagnosis and treatment. *Williams and Wilkins, Baltimore*,

Corthals, P. (2008). *Fonetiek en psycholinguïstiek* (Vol. Tweede, herziene uitgave: 2014-2015). ACCO.

Dacakis, G. (2000). Long-term maintenance of fundamental frequency increases in male-to-female transsexuals. *Journal of Voice*, *14*(4), 549–556. https://doi.org/10.1016/S0892-1997(00)80010-7

Dacakis, G., Davies, S., Oates, J. M., Douglas, J. M., & Johnston, J. R. (2013). Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals. *Journal of Voice*, *27*(3), 312–320.

Dacakis, G., Oates, J., & Douglas, J. (2012). Beyond voice: Perceptions of gender in male-to-female transsexuals. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *20*(3), 165–170. https://doi.org/10.1097/MOO.0b013e3283530f85

Davies, S., & Goldberg, J. M. (2006). Clinical aspects of transgender speech feminization and masculinization. *International Journal of Transgenderism*, *9*(3–4), 167–196.

Davies, S., Papp, V. G., & Antoni, C. (2015). Voice and communication change for gender nonconforming individuals: Giving voice to the person inside. *International Journal of Transgenderism*, *16*(3), 117–159.

De Bodt, M., Heylen, L., Mertens, F., Vanderwegen, J., & Van de Heyning, P. (2015). *Stemstoornissen. Handboek voor de klinische praktijk* (Zesde, Herziene Uitgave, Vol. 5). Maklu.

De Bodt, M., Patteeuw, T., & Versele, A. (2015). Temporal variables in voice therapy. *Journal of Voice*, *29*(5), 611–617.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning, P., Remacle, M., & Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology*, *258*(2), 77–82. https://doi.org/10.1007/s004050000299

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). Harper & Row.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 22–30.

Fitch, J. L. (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *Journal of Speech and Hearing Disorders*, *55*(2), 360–363. https://doi.org/10.1044/jshd.5502.360

Froeschels, E. (1952). Chewing method as therapy: A discussion with some philosophical conclusions. *AMA Archives of Otolaryngology*, *56*(4), 427–434. https://doi.org/10.1001/archotol.1952.00710020449010

Gallena, S. J. K., Stickels, B., & Stickels, E. (2018). Gender perception after raising vowel fundamental and formant frequencies: Considerations for oral resonance research. *Journal of Voice*, *32*(5), 592–601. https://doi.org/10.1016/j.jvoice.2017.06.023

Gelfer, M. P., & Tice, R. M. (2013). Perceptual and acoustic outcomes of voice therapy for male-to-female transgender individuals immediately after therapy and 15 months later. *Journal of Voice*, *27*(3), 335–347. https://doi.org/10.1016/j.jvoice.2012.07.009

Gelfer, M. P., & Van Dong, B. R. (2013). A preliminary study on the use of vocal function exercises to improve voice in male-to-female transgender clients. *Journal of Voice*, *27*(3), 321–334. https://doi.org/10.1016/j.jvoice.2012.07.008

Gooren, L. (2005). Hormone treatment of the adult transsexual patient. *Hormone Research in Paediatrics*, *64*(Suppl. 2), 31–36. https://doi.org/10.1159/000087751

Günzburger, D. (1995). Acoustic and perceptual implications of the transsexual voice. *Archives of Sexual Behavior*, *24*(3), 339–348.

Hancock, A. B., & Garabedian, L., M. (2013). Transgender voice and communication treatment: A retrospective chart review of 25 cases. *International Journal of Language & Communication Disorders*, *48*(1), 54–65. https://doi.org/10.1111/j.1460-6984.2012.00185.x

Hancock, A. B., & Helenius, L. (2012). Adolescent male-to-female transgender voice and communication therapy. *Journal of Communication Disorders*, *45*(5), 313–324. https://doi.org/10.1016/j.jcomdis.2012.06.008

Hancock, A. B., Krissinger, J., & Owen, K. (2011). Voice perceptions and quality of life of transgender people. *Journal of Voice*, *25*(5), 553–558. https://doi.org/10.1016/j.jvoice.2010.07.013

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Kalra, M. (1978). Voice therapy in the case of a transsexual. *British Journal of Sexual Medicine*, *5*(40), 47–50.

Kawitzky, D., & McAllister, T. (2020). The effect of formant biofeedback on the feminization of voice in transgender women. *Journal of Voice*, *34*(1), 53–67. https://doi.org/10.1016/j.jvoice.2018.07.017

Kaye, J., Bortz, M. A., & Tuomi, S. K. (1993). Evaluation of the effectiveness of voice therapy with a male-to-female transsexual subject. *Scandinavian Journal of Logopedics and Phoniatrics*, *18*(2–3), 105–109.

Kayikci, M. E. K., Akan, S., Ciger, S., & Ozkan, S. (2012). Effects of Hawley retainers on consonants and formant frequencies of vowels. *The Angle Orthodontist*, *82*(1), 14–21.

Kmet, L. M., Lee, R. C., Cook, L. S. (2004). Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Edmonton: Alberta Heritage Foundation for Medical Research (AHFMR). AHFMR - HTA Initiative #13. 2004.

Ladefoged, P. (1993). *A course in phonetics*. Harcourt Brace.

Leung, Y., Oates, J., & Chan, S. P. (2018). Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, *61*(2), 266–297. https://doi.org/10.1044/2017_JSLHR-S-17-0067

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice*, *24*(5), 540–555. https://doi.org/10.1016/j.jvoice.2008.12.014

McNeill, E. J., Wilson, J. A., Clark, S., & Deakin, J. (2008). Perception of voice in the transgender client. *Journal of Voice*, *22*(6), 727–733. https://doi.org/10.1016/j.jvoice.2006.12.010

Meerschman, I., Claeys, S., Bettens, K., Bruneel, L., D'haeseleer, E., & Van Lierde, K. (2019). Massed versus spaced practice in vocology: Effect of a short-term intensive voice therapy versus a long-term traditional voice therapy. *Journal of Speech, Language, and Hearing Research, 62*(3), 611–630.

Meerschman, I., Van Lierde, K., Ketels, J., Coppieters, C., Claeys, S., & D'haeseleer, E. (2019). Effect of three semi-occluded vocal tract therapy programmes on the phonation of patients with dysphonia: lip trill, water-resistance therapy and straw phonation. *International Journal of Language & Communication Disorders*, *54*(1), 50–61.

Meister, J., Kuhn, H., Shehata-Dieler, W., Hagen, R., & Kleinsasser, N. (2017). Perceptual analysis of the male-to-female transgender voice after glottoplasty-the telephone test. *The Laryngoscope*, *127*(4), 875–881. https://doi.org/10.1002/lary.26110

Mészáros, K., Vitéz, L. C., Szabolcs, I., Góth, M., Kovács, L., Görömbei, Z., & Hacki, T. (2005). Efficacy of conservative voice treatment in male-to-female transsexuals. *Folia Phoniatrica et Logopaedica*, *57*(2), 111–118.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*, 1. https://doi.org/10.1186/2046-4053-4-1

Mordaunt, M., Adler, R. K., Hirsch, S. (2006). Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide. *Plural Publishing*.

Mount, K. H., & Salmon, S. J. (1988). Changing the vocal characteristics of a postoperative transsexual patient: A longitudinal study. *Journal of Communication Disorders*, *21*(3), 229–238.

Nolan, I. T., Morrison, S. D., Arowojolu, O., Crowe, C. S., Massie, J. P., Adler, R. K., Chaiet, S. R., & Francis, D. O. (2019). The role of voice therapy and phonosurgery in transgender vocal feminization. *The Journal of Craniofacial Surgery*, *30*(5), 1368–1375. https://doi.org/10.1097/scs.0000000000005132

Nordström, P.-E. (1977). Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics*, *5*(1), 81–92. https://doi.org/10.1016/S0095-4470(19)31115-5

Oates, J. (2006). Evidence based practice in voice therapy for transgender/transsexual clients. Voice and Communication Therapy for the Transgender/transsexual Client: A Comprehensive Clinical Guide. 45–68.

Oates, J ., & Dacakis, G. (1983). Speech pathology considerations in the management of transsexualism–A review. *British Journal of Disorders of Communication*, *18*(3), 139–151.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*, 210. [PMC*][10.1186/s13643-016-0384-4]* [27919275*]*

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201–2212. https://doi.org/10.1121/1.3552866

Quinn, S., & Swain, N. (2018). Efficacy of intensive voice feminisation therapy in a transgender young offender. *Journal of Communication Disorders*, *72*, 1–15. https://doi.org/10.1016/j.jcomdis.2018.02.001

Sackett, D. L. (1989). Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*, *95*(2), 2S–4S. https://doi.org/10.1378/chest.95.2_Supplement.2S

Shriberg, L., & Kent, R. (2003). *Consonants in clinical phonetics*. Pearson Education Inc.

Söderpalm, E., Larsson, A., & Almquist, S. (2004). Evaluation of a consecutive group of transsexual individuals referred for vocal intervention in the west of Sweden. *Logopedics Phoniatrics Vocology*, *29*(1), 18–30. https://doi.org/10.1080/14015430310021618

Stemple, J. (1984). *Clinical voice pathology: Theory and management*. Charles E. Merrill Publishing, Co.

Timmermans, B. (2013). *Klink Klaar. Uitspraak- en intonatiegids voor het Nederlands*. Davidsfonds Uitgeverij.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*(4), 1699–1707. https://doi.org/10.1121/1.397959

Van Borsel, J., De Cuypere, G., Rubens, R., & Destaerke, B. (2000). Voice problems in female-to-male transsexuals. *International Journal of Language & Communication Disorders*, *35*(3), 427–442.

Verdolini Abbott, K. (2008). *Lessac-Madsen resonant voice therapy. Clinician* manual. Plural Publishing.

Weirich, M., & Simpson, A. P. (2018). Gender identity is indexed and perceived in speech. *PloS One*, *13*(12), e0209226.

Wu, K., & Childers, D. G. (1991). Gender recognition from speech. Part I: Coarse analysis. *The Journal of the Acoustical Society of America*, *90*(4), 1828–1840. https://doi.org/10.1121/1.401663

## Appendix

**Levels of evidence from** Sackett (1989)

| Level | Type of evidence |
|---|---|
| I | Large RCTs with clear cut results |
| II | Small RCTs with unclear results |
| III | Cohort and case-control studies |
| IV | Historical cohort or case-control studies |
| V | Case series, studies with no controls |